1 An Unbiased Coin

An unbiased coin X (which takes on values 0 and 1, each with probability 1/2 has mean

$$\mu = \mathbb{E}[X] = \frac{1}{2}0 + \frac{1}{2}1 = 1/2.$$

(We will always use μ for the mean of the random variables.) And it has variance

$$\text{Var}(X) = \mathbb{E}[(X-\mu)^2] = \frac{1}{2}(0-\frac{1}{2})^2 + \frac{1}{2}(1-1/2)^2 = \frac{1}{4}.$$

Another equivalent expression for variance is

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

But X takes on value 0 and 1, so $X^2 = X$. And so

$$\mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X] - \mathbb{E}[X]^2 = \frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

We will also use σ^2 for the variance, and σ for the standard deviation. Hence, an unbiased $\{0,1\}$ r.v. has $\mu = 1/2$ and $\sigma^2 = 1/4$.

Exercise #1: show that if the coin comes up heads with probability p, then $\mathbb{E}[X] = p$ and $\text{Var}(X) = p(1-p) \leq \mathbb{E}[X]$.

1.1 Sums of Random Variables

Linearity of expectation says:

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Linearity of expectations is true even for correlated random variables, which are not independent. For independent random variables, we get more—we can prove that variances add up. So

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \qquad \text{for independent } X, Y.$$

Exercise #2: suppose $X_1, X_2, \ldots, X_n$ all have the same mean μ and variance σ^2, and are all independent. If $S = \sum_{i=1}^n X_i$. Show that $\mathbb{E}[S] = n\mu$ and $\text{Var}(S) = n\sigma^2$. Use the calculations above to infer that if the sum S of n unbiased coin tosses has mean $\mu = n/2$ and variance $\sigma^2 = n/4$.

Exercise #2b: Let T be the sum of n coin tosses (where now the coin has bias $p = \frac{1}{\sqrt{n}}$). Use Exercise #1 to infer that T has mean $\mathbb{E}[T] = \sqrt{n}$ and variance $\text{Var}(T) = \sqrt{n}(1 - \frac{1}{\sqrt{n}}) = \sqrt{n} - 1$. Moreover, observe that $\text{Var}(T) \leq \mathbb{E}[T]$.

2 The Two Questions

We are interested in two questions:

1. Given a random variable X, and some value ℓ, what is the value of $\Pr[X \geq \ell]$? I.e., what is the likelihood of being in the "tail" of the distribution?
2. Given a random variable X, and some target likelihood δ (say $\delta = 1/n^2$ for concreteness). What is the value of ℓ_δ such that the likelihood of being in the "tail" of the distribution is at most δ?

Most of the time, we solve the former problem (for a general ℓ) to get some probability value that is a function of ℓ. Then we work backwards from there—set this probability value to δ and figure out what ℓ that corresponds to.

3 The Concentration Bounds

3.1 Markov's Inequality

Let X be a *positive* random variable. Then for all $k \geq 1$,

$$\Pr[\ X \geq k\mu\ ] \leq \frac{1}{k}. \tag{1}$$

Exercise #3: Let S be sum of n unbiased coin tosses. Show that $\Pr[S \geq 3n/4] \leq 2/3$.

Let T be as in Exercise #2b. Show that $\Pr[T \geq 3n/4] \leq \frac{4}{3\sqrt{n}}$.

3.2 Chebyshev's Inequality

For any r.v. X with mean μ and variance σ^2, the following holds.

$$\Pr[\ |X - \mu| \geq k\sigma\ ] \leq \frac{1}{k^2}.$$

Now we don't need X to be non-negative. This is *Chebyshev's inequality*. The proof, interestingly, just applies Markov's to the r.v. $Y = (X - E[X])^2$.

Exercise #4: Let S be sum of n unbiased coin tosses. Use Chebyshev to show that $\Pr[S \geq 3n/4] \leq 2/n$. Show that

$$\Pr[S \geq n/2 + t\sqrt{n}/2] \leq 1/t^2. \tag{2}$$

Observe that in order to a tail probability of $\delta = 1/n$, we would have to set $t = \sqrt{n}$. So we would only be able to say $\Pr[S \geq n/2 + n/2] \leq 1/n$, which is very weak. (In fact the probability is zero.)

Let T be as in Exercise #2b. Show that $\Pr[T \geq 3n/4] \leq O(\frac{1}{n^{3/2}})$.

3.3 Chernoff/Hoeffding Bounds

Now we'll assume more about the random variables: that they are sums of "small" random variables (say taking on values between 0 and 1) that are independent.

Theorem 1 (Hoeffding's Bound) *Suppose $X = X_1 + X_2 + \ldots + X_n$, where the X_is are independent random variables taking on values in the interval $[0,1]$. Let $\mu = E[X] = \sum_i E[X_i]$. Then*

$$\Pr[X > \mu + \lambda] \leq \exp\left(-\frac{\lambda^2}{2\mu + \lambda}\right) \tag{3}$$

$$\Pr[X < \mu - \lambda] \leq \exp\left(-\frac{\lambda^2}{3\mu}\right) \tag{4}$$

Exercise #4: Let S be sum of n unbiased coin tosses. Use Chernoff to show that

$$\Pr[S \geq 3n/4] \leq \exp\left(-\frac{(n/4)^2}{n + n/4}\right) = \exp(-n/20).$$

Show that for $t \leq 2\sqrt{n}$,

$$\Pr[S \geq n/2 + t\sqrt{n}/2] \leq \exp\left(-\frac{t^2 n/4}{n + t\sqrt{n}/2}\right) \leq \exp\left(-\frac{t^2 n/4}{2n}\right) = e^{-t^2/8}. \tag{5}$$

Observe that we're getting a much better bound than in (2), by using more properties about S, about it being a sum of small independent r.v.s.

Use this to show that if we want the tail probability $\Pr[S \geq \mu + \lambda]$ being at most $\delta = 1/n^2$, show that setting $\lambda = 2\sqrt{n \ln n}$ suffices. (Hint: if you want the right side of (5) to be $1/n^2$, what value of t would you set? What value of λ would that correspond to.

4 Union Bound

Very often we will have a collection of "bad" events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$. Then

$$\Pr[\text{at least one of the bad events } \mathcal{E}_i \text{ happens}] = \Pr[\cup_i \mathcal{E}_i] \leq \sum_i \Pr[\mathcal{E}_i].$$

If we want none of these n bad events to happen, then we can set $\Pr[\mathcal{E}_i] \leq 1/n^2$. Then

$$\Pr[\text{at least one of the bad events } \mathcal{E}_i \text{ happens}] \leq \sum_i (1/n^2) \leq 1/n.$$

So

$$\Pr[\text{none of the bad events } \mathcal{E}_i \text{ happens}] \geq 1 - 1/n.$$

5 Balls into Bins

This is the example from lecture. We throw n balls into n bins, uniformly and independently. Want to control the maximum load of any bin.

5.1 Defining the Random Variables

So if X_{ij} is the indicator for whether ball i goes into bin j, then $X_{ij} = 1$ with probability $1/n$, and 0 otherwise. Note that $X_{ij}, X_{ij'}$ are *not* independent of each other, since $X_{ij} = 1$ means $X_{ij'} = 0$ for $j' \neq j$, the ball i can only go into one bin. However $X_{1j}, X_{2j}, \ldots, X_{nj}$ are all independent, since the choices for the balls are made independently.

Let L_j be the load of bin j. Then

$$L_j = \sum_{i=1}^{n} X_{ij}. \tag{6}$$

And the right side is a sum of independent r.v.s. (You will almost surely use a Chernoff bound.) We want to control the max-load. That is

$$L_{\max} = \max_j L_j.$$

What result do we want? Today we want to show that

$$\Pr[L_{\max} \le \text{blah}] \ge \text{large} = 1 - 1/n. \tag{7}$$

This is a "with high probability" result.

The max of a bunch of random variables is not as well-behaved as a sum. Hence, to show (7), we often use a union bound—show that each of the L_j variables (one for each bin) are smaller than *blah* with probability $1/n^2$, and then take a union bound over all bins j.

5.2 Bounding Load of Each Bin

Let's use B for blah. Fix a bin j, and we want to find B such that

$$\Pr[L_j \ge B] \le 1/n^2. \tag{8}$$

Let's see what we can do with our favorite concentration bounds above:

5.2.1 Prelim Calculations

What is the expectation $\mathbb{E}[X_{ij}]$? It is

$$(1 - \frac{1}{n}) \cdot 0 + \frac{1}{n} \cdot 1 = 1/n.$$

Hence by linearity of expectations:

$$\mathbb{E}[L_j] = \mathbb{E}\left[\sum_i X_{ij}\right] = \sum_i \mathbb{E}[X_{ij}] = n \cdot 1/n = 1.$$

Note that if we had m balls, the expected load would be m/n. As expected (har har).

What's the variance? You can calculate it by hand, but not that each X_{ij} is like a coin with bias $p = 1/n$, so you can read off its variance from Section #1 as being $p(1-p) = \frac{n-1}{n^2}$. And since L_i is a sum of independent coin tosses, their variance adds up to give $\text{Var}(L_j) = n \cdot \frac{n-1}{n^2} = \frac{n-1}{n}$.

5.2.2 Markov

The mean of L_j is $\mu = 1$, and we want to find B such that the tail probability is $1/n^2$. So all Markov can say is that we will not be more than $n^2\mu$ with probability $1/n^2$. Useless.

5.2.3 Chebyshev

Again, we want to prove (8) for as small a B as possible, maybe using Chebyshev. So let's get (8) into a Chebyshev-like form. (Recall $\mu = \mathbb{E}[L_j] = 1$.)

$$\Pr[L_j \geq B] = \Pr[L_j - \mu \geq (B-1)] \leq \Pr\left[|L_j - \mu| \geq \Big(\frac{B-1}{\sigma}\Big)\sigma\right].$$

Chebyshev says that deviating by k times σ has likelihood at most $1/k^2$. So deviating by $(B-1)/\sigma$ times σ has likelihood at most $\sigma^2/(B-1)^2$. Set this equal to $1/n^2$, what we want. You get

$$\frac{\sigma}{B-1} = \frac{1}{n} \iff B = 1 + \sigma n = 1 + (n-1) = n.$$

So Chebyshev says: the chance of the load being more than $B = n$ is at most $1/n^2$.

Again useless, since we know that no load can be more than n anyways, there are only n balls.

5.2.4 Chernoff

Now to prove (8) for as small a B as possible using Chernoff. Again, get it into a Chernoff-like form:

$$\Pr[L_j \geq B] = \Pr[L_j \geq \mu + (B-1)].$$

This is at most $\exp(-\frac{(B-1)^2}{2\mu+B-1})$. So set this equal to $1/n^2$ and solve for B.

$$\exp(-\frac{(B-1)^2}{2\mu + B - 1}) = \frac{1}{n^2} \iff \exp(\frac{(B-1)^2}{2\mu + B - 1}) = n^2.$$

Take logs to get

$$\frac{(B-1)^2}{2\mu + B - 1} = 2\ln n.$$

But $\mu = 1$, so

$$\frac{(B-1)^2}{B+1} = 2\ln n.$$

Now solving for B exactly is annoying. But if we choose B slightly larger than strictly necessary, we'd still be fine, the probability would be even lower.

Here's what I'd do now: $B-1$ and $B+1$ are close to B (if B is large). So squint, then the LHS becomes like $B^2/B = B$. So we want $B \approx 2\ln n$. Probably we lost a bit in this approximation, so try $B = 16\ln n$. Then

$$\frac{(B-1)^2}{B+1} \geq \frac{(B/2)^2}{2B} = \frac{B}{8} = 2\ln n.$$

As desired.[1]

[1] Of course, we could have been much more careful with these approximations, but that's not the point—for now. We just want a ballpark estimate, which we can refine at leisure.

5.2.5 Wrapup

So we showed: the load L_j of each machine j is more than $B = 16 \ln n$ with probability $\leq 1/n^2$. Union bounding over all these "bad" events, the probability that *some* machine having load more than B is $1/n$. So the probability of all machines have load at most $O(\ln n)$ is $1 - 1/n$. This proves (7).